

Date: June 19, 2009
From: Clifton Bell, Malcolm Pirnie
Re: Review of CFD and Reference Curve Revisions

The Virginia and Maryland Associations of Municipal Wastewater Agencies (V/MAMWA) have provided technical comments on the cumulative distribution frequency distribution (CFD) assessment approach during several prior public comment periods. For example, V/MAMWA submitted extensive comments on the draft 2006 criteria addendum, including discussion of the bias in the CFD method towards non-attainment. As such, it is a positive development that the Chesapeake Bay Program (CBP) is recognizing the bias issue and attempting to address it. Given the regulatory and economic repercussions of 303(d)-listing, there are few if any technical issues that are more important to address.

On behalf of V/MAMWA, Malcolm Pirnie has reviewed the current efforts to lower and balance the false positive/negative assessment errors. In general, the percentile-based approach to revising the reference curves appears to be technically defensible if it can achieve low and balanced rates of false positive/negative findings of non-attainment. We offer the following technical comments on this issue:

1. *It is recommended that the CBP examine the sensitivity of the reference curves to the data screening criteria:* For the current effort, reference segment-periods were selected not only by the average B-IBI score (≥ 3.0), but also by a minimum B-IBI sample size ($n \geq 10$) and a maximum B-IBI standard deviation ($s < 1.0$). V/MAWA supports the concept of screening segment-periods based on n and variability. However, the minimum n and standard deviation are partly arbitrary. The present analysis has result in a relatively small number of reference segment-periods for consideration. If slightly different data screening criteria resulted in a larger number of segment-periods for evaluation, they should be considered.

To explore this issue, Malcolm Pirnie investigated how many more deep water segment-periods would be included if the average B-IBI criterion were maintained at 3.0, but the minimum n requirement was changed from 10 to 8 (Scenario B) and/or the maximum standard deviation was changed from 1.0 to 1.2 (Scenario C). Results (Table 1) indicate that slight alteration of the data screening criteria could substantially increase the number of “healthy” segment-periods for analysis.

The 100th percentile reference curve can be sensitive to the inclusion of even one additional segment-period. For example, the segment-period CB5MH 1999-2001 has an average B-IBI of 3.2, n of 10, and standard deviation of 1.08. Because it (just barely) misses the standard deviation, it would be excluded from the reference curve analysis

under CBP's default data screening criteria. Yet if included, this segment would be the "controlling" station for a portion of the 100th percentile reference curve.

Given the fact that the 100th percentile-based reference curve can be sensitive to the inclusion of even one additional station, and that the data screening criteria are partially arbitrary, it is recommended that the CBP explore the sensitivity of the reference curve to the data screening criteria, and include as many segment-periods as can be technically justified.

TABLE 1
Number of "Healthy" Deep Water Segment Periods Based on B-IBI Database Screening Criteria

[Based on 1996-2007 data; each segment-period represents a single segment over a 3-year assessment period]

	Scenario A (Default) B-IBI ≥ 3.0 $n \geq 10$ S.D. < 1.0	Scenario B B-IBI ≥ 3.0 $n \geq 10$ S.D. < 1.2	Scenario C B-IBI ≥ 3.0 $n \geq 8$ S.D. < 1.0	Scenario D B-IBI ≥ 3.0 $n \geq 8$ S.D. < 1.2
Total number of "healthy" deep water segment- periods	15	16	22	26

2. The new data screening criteria used to developed reference curves might lead to underestimation of assessment error rates: As discussed in the last WQSC face-to-face meeting, the published reference curve results in a very high rate of false positive findings of non-attainment. The percentile-based reference curves under development purportedly have much lower rates of false positives. However, much of this effect is not necessarily caused by a more accurate reference curve (after all, the published and draft revised curves are very similar) but by the new data screening procedures that remove from the analysis many of the segment-periods that were previously found to be false positives. As shown in Table 1, there are numerous segment-periods with "healthy" B-IBI scores that were not included in the most recent analysis. Although we did not have the DO violation rates for all of these segment-periods, at least some of them (*e.g.*, CB5MH 1999-2001, which had a B-IBI score of 3.2) would show up as false positives if compared to the proposed 100th-percentile-based reference curve for deep water.

Similarly, there are segment-periods with B-IBI scores less than 3.0 that were not classified as "degraded" due to having fewer than 10 B-IBI observations. If a slightly different criterion (*e.g.*, n of at least 8) were used, some of these segments could rightly be considered degraded segments. Depending on what percentile is chosen to serve as the basis of the reference curve, exclusion of these segments could cause underestimation of the rate of false negatives.

The possibility of underestimating error rates is offered as another reason to closely evaluate the data screening criteria used to develop the reference curves, and to ensure that healthy or degraded segment-periods are not being excluded from the analysis. Relatedly, the evaluation of false positive/negative assessment rates should consider the error rates associated with segment-periods that would have been included if slightly different data screening criteria were used.

3. Evaluation of the area under the assessment curve represents a viable alternative method for balancing false positives/negatives: As V/MAMWA pointed out in their 2006 comments, the present assessment method's bias toward non-attainment is largely driven by that fact that non-attainment is determined by exceedance of the reference curve at any temporal axis point, without regard to area where the assessment curve is below the reference curve. A segment-period is out of compliance for the slightest exceedance but does not get any credit for being better than reference conditions in other parts of the curve. The issue is especially problematic near the "horns" of the assessment curve (*i.e.*, in the upper left and lower right part of the curve), where the reference and assessment curves will converge by geometric necessity. A segment can have significantly better overall water quality than reference conditions but still fail the assessment.

Given the very high variability in the shape of the CFD curves from both reference and non-reference segments, the data do not allow the identification of highly specific combinations of time-and volume that lead to poor B-IBI scores in a segment. Rather, low B-IBI scores can be more confidently linked to the overall rate of exceedance in time-space than to subtleties in the shape of the curves. Therefore, the solution to the non-attainment bias problem offered by V/MAMWA in 2006 still merits consideration today: retain the CFD approach, but determine attainment based on the total area under the assessment curve. This area would be compared to the area under the reference curve to determine attainment (Figure 1).

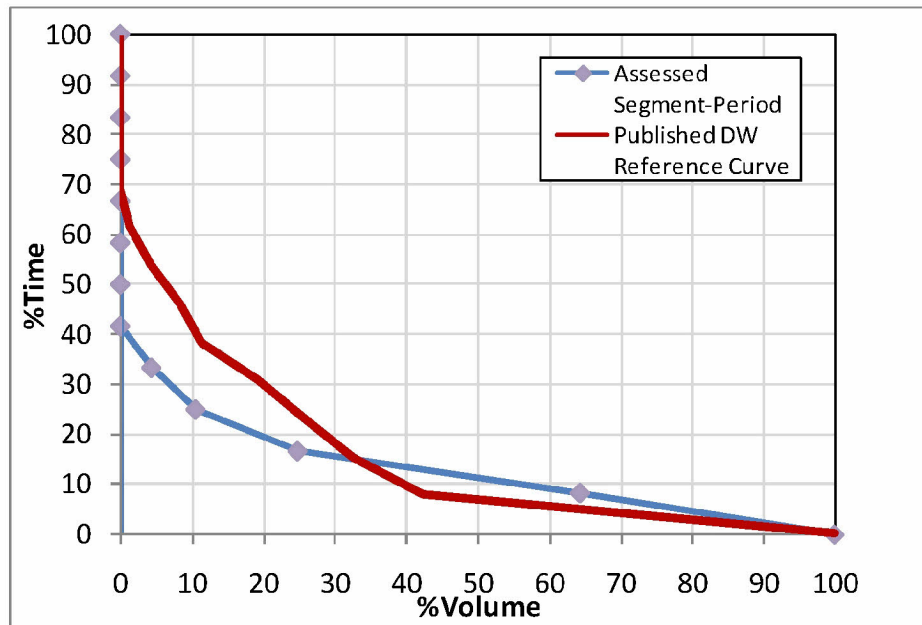


Figure 1: Example of a segment period that would fail under the “point” approach but pass under the “area” approach.

An area-under-the-curve approach would retain the basic intent of the CFD approach, which is the joint consideration of temporal and spatial violation rates relative to a reference condition. However, it would be inherently more balanced with regard to false positive/negative findings of impairment. To explore this, Malcolm Pirnie compared the two methods when applied to a list of deep water healthy and degraded segment-periods recently determined by the CBP. Each segment was assessed using the published “point” method and the proposed “area” method, both using the published deep water reference curve. Areas under the curve were estimated using the trapezoidal method.

Results (Tables 2-3) demonstrate that the area method has a more balanced and overall lower error rate than the “point” method. The actual error rates are expected to change as the data screening requirements and other methods for defining reference conditions change. However, the area method is expected to retain its superior balance of false positives/negatives.

An area-based assessment method would be especially beneficial for assessment of criteria that use the default 10% reference curve. This curve has no specific biological basis, and was derived as a replacement to the older, equally arbitrary “10% rule.” But unlike the older “10% rule,” the point method causes the reference curve to have an inherent bias towards false positives, as discussed above. For criteria that use a default 10% reference curve, the area method may provide the only means to balance false positive/negative error rate.

TABLE 2
Comparison of Correct and Incorrect Assessments Using the
“Point” and “Area” Methods

[Based on “healthy” and “degraded” segment list provided by CBP, 5/14/2009]

Method	Correct		Incorrect	
	Healthy Segments Passing	Degraded Segments Failing	Healthy Segments Failing	Degraded Segments Passing
Published “Point” Method	64%	100%	36%	0%
Proposed “Area” Method	100%	89%	0%	11%

TABLE 3
Comparison of Correct and Incorrect Assessed Segment Periods Using the
“Point” and “Area” Methods

[Based on “healthy” and “degraded” segment list provided by CBP, 5/14/2009]

Method	Correct		Incorrect	
	Healthy Segments Passing	Degraded Segments Failing	Healthy Segments Failing	Degraded Segments Passing
Published “Point” Method	CB6PH19961998 CB7PH19961998 CB7PH19971999 CB7PH19982000 CB7PH19992001 CB7PH20002002 CB7PH20042006	POTMH19992001 POTMH19982000 PATMH20042006 PATMH20032005 RPPMH20022004 PAXMH19992001 PAXMH20012003 PAXMH20042006 POTMH19971999 PAXMH20032005 PAXMH20002002 POTMH19961998 RPPMH20002002 PAXMH20022004 PAXMH19961998 PAXMH19971999 YRKPH20042006 YRKPH20032005	CB6PH19971999 CB6PH19992001 CB6PH20042006 YRKPH20022004	
Proposed “Area” Method	CB6PH19961998 CB6PH19971999 CB6PH19992001 CB6PH20042006 CB7PH19961998 CB7PH19971999 CB7PH19982000 CB7PH19992001 CB7PH20002002 CB7PH20042006 YRKPH20022004	POTMH19992001 POTMH19982000 PATMH20042006 PATMH20032005 RPPMH20022004 PAXMH19992001 PAXMH20012003 PAXMH20042006 POTMH19971999 PAXMH20032005 PAXMH20002002 POTMH19961998 RPPMH20002002 PAXMH20022004 PAXMH19961998 PAXMH19971999		YRKPH20042006 YRKPH20032005

4. For open water, CBPO should consider use of the 78th-percentile-based reference curve until an alternative biological reference curve can be developed. It is our understanding that the CBPO is considering use of the default 10% reference curve for open water, on the basis that the B-IBI is not an appropriate measure of open water biological reference conditions. However, the default 10% curve has an even weaker biological linkage than the B-IBI derived curves. It is recommended that the CBPO

explore alternative means to define biological reference conditions for open water. Until that time, the B-IBI based curves represent the strongest biological linkage available. The 78th-percentile-based reference curve has the best balance between false positive and negative findings of B-IBI non-attainment, and therefore should receive first consideration for use in assessment.

cfb